# Performance evaluation of road detection and tracking algorithms

D. Dufourd and A. Dalgalarrondo
DGA/Centre Technique d'Arcueil
16bis, av. Prieur de la Côte d'Or, 94114 Arcueil Cedex, France,
Delphine.Dufourd@etca.fr, Andre.Dalgalarrondo@etca.fr

## ABSTRACT

In this paper, we present a methodology to assess the results of image processing algorithms for unstructured road edges detection and tracking. We aim at performing a quantitative, comparative and repetitive evaluation of numerous algorithms in order to direct our future developments in navigation algorithms for military unmanned vehicles. The main scope of this paper is the constitution of this database and the definition of the assessment metrics.

**KEYWORDS:** *Image Processing Assessment, Outdoor Navigation, Ground Robotics.*

## 1 GOAL OF OUR WORK

In December 1999, the French defence procurement agency (Délégation Générale pour l'Armement) has launched a prospective program dedicated to ground robotics. Part of this program aims at developing autonomous functions for military unmanned vehicles navigation, such as autonomous road following, beacon and vehicle tracking and scene analysis. In this context, the Centre Technique d'Arcueil (CTA) of the DGA is currently conducting an evaluation of existing image processing detectors of unstructured road edges. The goal of this evaluation is to compare different road detection and following algorithms in a reproducible and quantitative way so as to direct future developments in navigation algorithms. It should allow us to determine the most promising techniques and possibly find orthogonal strengths between the algorithms so as to conceive hybrid and potentially more efficient methods. In this work, we plan to evaluate six road edges detectors coming from: Centre de Morphologie Mathématique (CMM) of the École des Mines de Paris [3], Laboratoire des Sciences et Matériaux pour l'Électronique et l'Automatique (LASMEA) [22, 1], Laboratoire Central des Ponts et Chaussées (LCPC) [7], the PG:ES company [23], and our laboratory [20].

The evaluation methodology is described in the following sections. Section 2 presents previous studies on performance evaluation. Section 3 focuses on our evaluation software environment named SENA. Section 4 describes the constitution of the image data base as well as the associated ground truth. Section 5 proposes different metrics to evaluate the algorithms with respect to the ground truth. Section 6 shows preliminary results concerning two road-following algorithms. Finally, section 7 concludes and outlines future developments.

## 2 EVALUATION METHODOLOGIES

### 2.1 *Assessment methods in image processing*

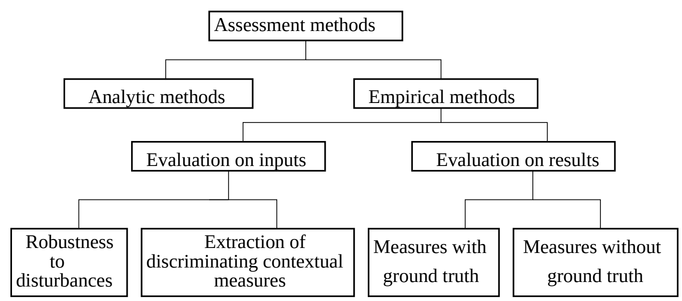


Figure 1: Classification of assessment methods.

In the last years, the image processing community has started to develop evaluation methods in order to be able to compare quantitatively the huge number of algorithms available after these last decades of research. Such an approach is very important for those who use image processing as a part of their research, like roboticists, since it provides a guide based on performance among the overwhelming available algorithms. However it should be noted that such an approach is very recent. For instance, Heath [12] has analyzed 21 papers on new contour detectors during the years 1993-96; the results are rather startling: while some papers do not even compare their method with other detectors, other papers use only 2 test images. Up to very recently, algorithms were not evaluated quantitatively, but only qualitatively on various criteria such as the neatness of their design or the sophistication of the underlying mathematical theoretical tools. Most experiments are conducted by human experts and lack any automation. The performance of the algorithm depends then on the

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

| | | |
|---|---|---|
| 1. REPORT DATE<br>**AUG 2002** | 2. REPORT TYPE | 3. DATES COVERED<br>**00-00-2002 to 00-00-2002** |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>**Performance evaluation of road detection and tracking algorithms** | 5a. CONTRACT NUMBER |
| | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**DGA/Centre Technique d'Arcueil,16bis, av. Prieur de la Cote d'Or,94114 Arcueil Cedex, France,** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**Proceedings of the 2002 Performance Metrics for Intelligent Systems Workshop (PerMIS ?02), Gaithersburg, MD on August 13-15, 2002**

14. ABSTRACT
**see report**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **8** | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

know-how and the personal experience of the expert. Fortunately, the situation is changing, following the animated discussion of Jain and Binford [16], and there are always more special issues in journals or conferences focusing on image processing assessment issues.

Figure 1, taken from [6], shows a temptative general classification of methods for image processing assessment.

Analytic methods do not need an explicit implementation of the algorithm and take into account its general features such as its complexity, or the overall principles. Such methods can be used in the development phase when the designer has to choose which algorithms will be implemented on the robot. They allow a comparison of the algorithmic complexity and give an estimate of the time to be allotted to every algorithm, when the computing resources are known. The influence of the propagation of the variance of the input data on the results of the algorithm can also be estimated [11].

Empirical methods evaluate the algorithm by playing with its inputs and studying the evolution of its various outputs. The assessment of an algorithm can be done by varying the intrinsic parameters of the algorithm or by adding disturbances – noise, time-depending variation of the grey levels, saturation...– on the inputs and analyzing the evolution of the performance. Such an approach aims at defining the "satisfactory operating domain" of the algorithm. Such a knowledge is important in order not only to compare and choose the right algorithm but also to chain various algorithms, as it gives hints at the propagation of errors. A weak sensitivity to disturbances or modification of the tuning parameters is needed in an automatic system. Some methods use contextual hopefully discriminating measures in order to decide whether an input – in our case the current image – "suits" the algorithm, i.e. is in the "satisfactory operating domain". Measures that are correlated with the result of the algorithm are looked for.

Methods based on the measure of a difference between the results of an algorithm and a reference solution, called "ground truth", allow an automation of the assessment process. As shown in figure 2, the joint use of test images, ground truth and metrics, that yield a measure of the difference between the results and the ground truth, provides quantitative evaluation of the algorithms. Whereas the ground truth is generated by a human expert or by a reference algorithm, the variation of the tuning parameters of the algorithm follows predetermined ranges and sampling and can be fully automated, as well as the analysis of the results, as soon as the metrics have been explicitly given. This is the method we have selected for our assessment.

Finally, empirical evaluation methods without ground truth are based on the availability of empirical measures of what a "correct result" should be [6]. Such measures are built following intuition and/or successive experiments during the design phase, where ground truths may be used. Of course such measures are very dependent on the task to be performed by the algorithm but they can be automatized. For example, in [19], we present a robot control architecture which uses evaluation mechanisms in order to select automatically the most appropriate perception algorithm.

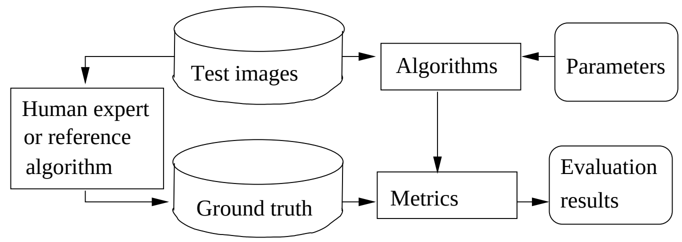


Figure 2: How to assess an algorithm when ground truth is available.

## 2.2 *Road following algorithms evaluation*

Although a wide variety of vision-based road following algorithms have been proposed and implemented over the last two decades, few techniques have been developed to assess their quality. Far too many articles rely on qualitative results, exhibiting a handful of example images to illustrate the performance of the algorithms while real applications would mean processing millions of images without making any serious error [17].

In many cases, the efficiency of road following algorithms is only characterized by the speed achieved by the whole autonomous system. For instance, in the field of autonomous lateral control on highways and marked roads, numerous experiments consist in driving a few thousands of kilometers and providing statistics about the performance of the system : maximum time elapsed between two manual interventions, average and maximum speed, distance between the vehicle and the lane, etc [5]. However, using such global characterizations, it seems difficult to determine exactly what makes the system efficient and what could be improved to make it better : is the autonomous vehicle fast because the road following algorithm has been implemented efficiently using powerful computationnal resources, because this image processing algorithm is very accurate and robust or because the control laws of the vehicle are well-designed?

Algorithms performing 3D reconstruction of the road have been evaluated in different ways. Guiducci performed indirect numerical tests on 1000 images, comparing the road width and vehicle speed estimated by his algorithm with their real values [10]. The actual road width was measured manually and the speed was given by the vehicle speedometer. However, these global test measures characterize the whole system, including the 3D road and vehicle models, while more direct measures would probably be helpful to improve the image processing algorithms more specifically. DeMenthon performed tests on both synthetic and real images [8]. Whereas the 3D profile of the synthetic data is known, the profile for the real data is reconstructed manually using a fusion between distance

and video images. A specific task-oriented metric is used to assess the results of the algorithm: a reconstructed road is labelled "navigable" if the tracks of a two meter-wide vehicle following the centerline of the reconstructed road stay between the edges of the actual road over the whole reconstruction and do not cut these edges. However, manual 3D reconstruction is too time-consuming if the evaluation is to be performed on numerous data. Therefore, if a manual ground truth is to be used, it seems more realistic to operate directly in the 2D image space rather than in the real 3D world.

Finally, a few research studies focus on automating the measurement of ground truth for the evaluation of vision-based lane sensing. A NIST report on performance evaluation for robotic vehicles [13] proposed a specific device composed of a side-looking camera and a separate vision system to measure the offset between the vehicle and the lane. Using a detailed calibration of their imaging system and spectral measurement of the ambient illumination and scene, Everson et al. [9] generated images simulating various rates of precipitation. The metric used to evaluate their lane-sensing system consists of the variance lane centering behavior as a function of precipitation level. Kluge also performed a pilot study in order to get some insight into the issues involved in automatic performance evaluation of lane-sensing algorithms [17]. He selected a well-defined aspect of system performance in a single class of lane-sensing techniques. The ground truth was measured automatically using a reference algorithm and its correctness was hand checked on the 1800 windows of the data set. One can notice that automatic ground truth measurement requires a reference algorithm and possibly a specific equipment to measure the road edges, which is easier in the case of road marking detection than in the case of unstructured road edges detection with various environmental conditions.

## 3 THE SENA PLATFORM

Our laboratory is interested in various information and intelligence military systems which use image processing. Current researches address the evaluation of satellite image registration, infrared image segmentation, image fusion and interpretation. Applications of these algorithms on military systems must present specific qualities in order to cope with extreme battlefield situations. This leads to different system testings and notably to the development of a general evaluation architecture called SENA (Système pour l'EvaluatioN d'Algorithmes).

SENA is a customized software environment for fast algorithm implementation and evaluation of a wide range of applications. It helps in assembling image processing operators and replaying the experiments on a large amount of images. In a sequence of operators, tools for measuring or visualizing partial results can be incorporated. These tools can also be considered as image processing operators. Thus, SENA is able to organize and execute a sequence of operators of different types (source code, shell scripts, binaries, libraries) and origins (operators that were developed specifically or not for the platform). The only constraint is that all the operators must be executed on the same host computer. Practically, SENA runs on a SMP computer (SUN Enterprise 10,000 with 32 processors) to cope with huge amounts of data and important range variation of the algorithms parameters. SENA has been developed by Cril Ingenierie under CTA specification and supervision. Among other graphical software environments able to construct and execute sequences of image processing operators, Khoros is probably the best known. However, SENA is most likely the only platform allowing simultaneous use of various types of operators (scripts, binaries...), definition of cyclic graphs of operators, automatic parallel execution of the assessment process on range of data and parameters and coupling with a database.

## 4 DATABASE CONSTITUTION

The database includes the images that will compose the input of the image processing algorithms and the ground truth suited to the final task to assess. For our purpose, we need images of unstructured roads and trails taken from a ground vehicle whose size and mobility are close to the targeted UGV. Collecting these images is relatively easy and cheap with nowadays technologies. The two main difficulties are the representativity of the images, in relation with the missions and the environment of the UGV, and the constitution of the ground truth.

The first step is the specification of the hardware to grab images on the proving ground. This includes the vehicle, the camera (position, field of view, frame rate, resolution, type of sensor...), the grabbing device, the storage media and the image files coding. Specification of noise and saturation levels on the images and general ranges of climatic or illumination conditions can be added. If image calibration is needed by some algorithms, the acquisition of images of reference scenes must be specified. Moreover, data concerning the speed and the attitude of the vehicle can be attached to each image, in order to feed the environment or vehicle models which may be used by some algorithms.

The second step is the specification of film scripts for the image acquisition. In our case, we specify two kinds of scenarios: general ones with an increasing difficulty level for road edges extraction and special scenarios which are dedicated to road and trail particularities. In the first case, one gets homogeneous sequences of images in order to assess an algorithm all along a sequence without risking an irreparable failure on some images. In the second case, it is possible to evaluate the algoritm behavior in harsh conditions. The special scenarios must provide known difficulties for the algorithms like puddles, hairpin bend, abrupt road widening, slough, parked vehicles on the roadsides, changing soil, transversal and longitudinal road markings, etc. In practice, we defined six general sce-

narios and twelve special scenarios. The general scenarios belong to two categories: tarmac roads and gravel-mud roads. There are three scenarios for each category, with an increasing level of difficulty. Each scenario must correspond to a specific location on the proving ground in order to be recorded in about four different illumination and weather conditions. Some image sequences recorded at night with the vehicle lights and with a FLIR camera are also defined. The length of the image sequences may vary between 60 and 120 s which corresponds to a distance between 500 and 1000 m for a vehicle travelling at a mean speed of 30 km/h. As for the twelve special scenarios, the length of the image sequences is shorter (about 20 to 30s) in order to isolate each difficulty.

The image acquisition is currently being performed in DGA testing facilities situated near Angers. Figure 3 shows examples of images taken at this location. This first version of the database will count about 20,000 images of roads and trails. This amount accounts for the second main difficulty of the construction of this database. Indeed, on each image a human expert has to define the ground truth i.e. to draw the road edges on the images. For that particular task, we wrote a specification which contains rules to follow in order to decide where the road edges are in a given image. Then, in order to facilitate this long and dull job, we have created a program with a dedicated graphical interface which manages the name and numbering convention of the images and ground truth files of a sequence and allows, on a new image, an easy modification of the grountruth defined on the previous image.

# 5 EVALUATION METRICS

Hoover et al. [14] underlined the need for multiple metrics in image processing algorithms assessment, so that users can consider different aspects of the algorithms and choose the one which is best suited to their application. Following this point of view, we propose eight different metrics aiming at assessing geometrical accuracy as well as a good global correspondence between the ground truth and the output of the algorithms. As mentionned before, extracting 3D references from numerous data appears extremely time-consuming so that we have decided to work in the 2D image space. Therefore, our metrics are also computed in the 2D image space and do not consider the 3D real world data such as the width of the road or the pitch angle of the vehicle.

Among the various metrics available, we can distinguish contour-oriented metrics and region-oriented metrics, which reflect the dual approaches to image segmentation.

## 5.1 *Contour-oriented metrics*

Before computing most contour-oriented metrics, we need to perform a matching procedure between the reference road edges and the result of the algorithm. Indeed, we have to determine which parts of the extracted road edges correspond to given parts of the reference road eges. We chose the so-called "buffer method" described by Wiedmann et al. [25] in the context of automatic road axes extraction from aerial images. Using this technique, every portion of the extracted road boundary lying within a certain distance (i.e. the size of the buffer) from the reference boundary is considered as matched.

A survey realized in our lab by Capolunghi and Ropert [6] defines five different categories for common assessment measures, as listed below.

### 5.1.1 Measures of classification/detection errors

These measures consist in counting the number of pixels that have been misclassified by the algorithm and extracting detection and cover rates as well as statistical measures. Our first three metrics correspond to this category.

The **completeness metric** computes the difference between the length of a result judged as valid (within the buffer tolerance) and the length of the ground truth. It enables us to determine whether the algorithm has managed to find the whole road or only a small part of it. More formally, using the notations and configuration of Fig. 4, this metric is defined by:

$$m_1 = \frac{length(BC)}{length(AD)}, \quad m_1 \in [0,1]$$

The **correction metric** determines what portion of the result lies within the tolerance area. Using the notations and configuration of Fig. 4, it is defined by the following formula:

$$m_2 = \frac{length(GF)}{length(GE)}, \quad m_2 \in [0,1]$$

Finally, a **quality metric** combines the previous ones. The quality of a road edge estimated by the algorithm is regarded as good if the edge lies within the tolerance area and "explains" most of the reference edge. More precisely, this quality can be expressed as: $m_3 = m_1 \times m_2, \quad m_3 \in [0,1]$.

Wiedmann et al. defined similar metrics using notions of true positive, false positive and false negative for the output of the algorithms [25].

### 5.1.2 Measures of localization errors

Measures of localization errors compute a distance between two sets of points $A$ and $B$ (in the case of contours, one can consider that these sets are composed of the pixels that form the contour). Among them, we can mention the figure of merite proposed by Pratt [21], the Hausdorff distance and the Baddeley distance [2] . Huang and Dom [15] also proposed to evaluate the divergence between $A$ and $B$ by distance distribution signatures which correspond to distance histograms. Different statistics can be extracted

Figure 3: Examples of road images of the DGA testing facilities near Angers.

from these histograms such as the mean value and variance. We have opted for this last measure computing the average distance between the reference and result edges:

$$m_4 = \frac{\Sigma_G^E dist(algorithm, ground\,truth)}{length(GE)}, \quad m_4 \in [0, \infty[$$

Besides, we can compute other statistics concerning these distances such as variance, as well as maximal and minimal distances (which is akin to the Hausdorff distance).

#### 5.1.3 Error classification

The evaluation of edge detectors is sometimes based on a classification of their errors. For example, the estimated edges can be labelled as well-detected contours, over- or under-segmented contours, missed contours and contours due to noise. Completeness and correctness illustrate some of these notions but we could also introduce new metrics involving "redundancy" [25] for instance. Fig. 4 illustrates this notion of redundancy: the length of the thick result edge far exceeds the length of the reference edge. However, the ground truth does not present many singularities and the algorithm results are usually smoothed by linear regressions or hyperbolic approximations, so that this redundancy metric potentially does not provide much information.

#### 5.1.4 Parametric approach

The measure classes described so far are computed pixel by pixel from the output data. Conversely, the parametric approach consists in representing the data which we intend to compare by a few specific features. As a result, the output data are reduced to a single parameter vector. For instance, Strickland proposed linear combinations of local measures related to the shape of the contour (continuity, regularity and thickness), its location with respect to the ground truth and to contours due to noise [24]. The first three criteria are not well-adapted to our application since the contours provided by the evaluated algorithms are usually one-pixel thick, continuous and regular. The fourth criterion related to location is taken into account by the measures of localization errors described above. However, the last criterion is not used because we have not measured the noise levels in the images.

#### 5.1.5 Non-scalar measures

Non-scalar measures can be linked to statistical approaches. For instance, Huang and Dom proposed distance histograms [15]. However, to avoid multiplying the measure data, we only keep the mean value for the histogram, and possibly the variance as well as the extreme values. Performance diagrams such as the Receiver Operating Characteristic (ROC) curves are often used to illustrate algorithms performance. We can draw similar curves representing $1 - m_2$ (which corresponds to a false positive rate) with respect to $1 - m_1$ (corresponding to a false negative rate) for different values of the algorithm parameters.

### 5.2 *Region-oriented metrics*

Contour-oriented metrics provide detailed information about the geometric accuracy of the algorithms. However, in sharp turns or on very irregular paths, pessimistic algorithms using a very simple road model (a triangle for example) risk being severely penalized by these metrics even if they find a drivable area within the boundaries of the real road. As a result, we have defined several metrics based on surfaces.

Whereas edge-oriented metrics need a preliminary matching process, region-based metrics can be applied directly since there is no ambiguity concerning their correspondence. However, in the general case, the road detectors provide open contours for the road, which means that we need to perform a closing procedure. We have decided to close the road region through linking the left and right upper ends as well as the left and right lower ends.

Region-oriented metrics can be divided into the same categories as contour metrics.

#### 5.2.1 Measures of classification/detection errors

Among the metrics measuring frequencies of incorrect classification of pixels in the image, we can mention the Hamming distance [15] and the Vinet distance. However, such

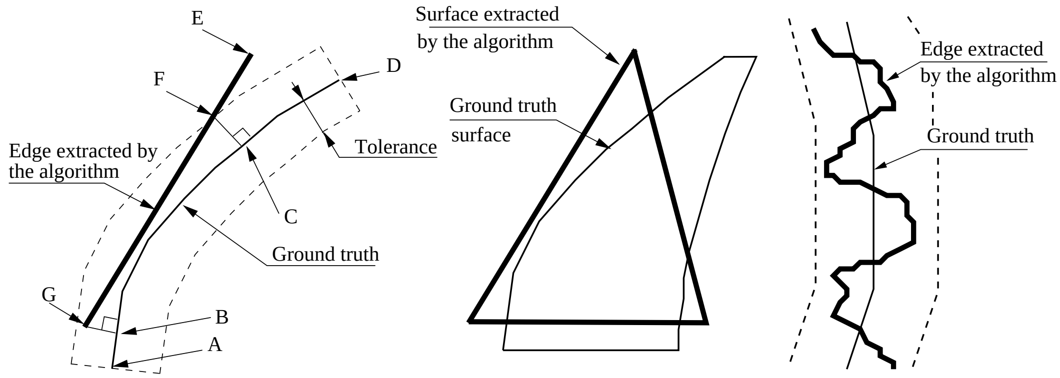


Figure 4: (left and center) Notations for metrics. (right) An example of redundancy.

metrics are designed to deal with region segmentation algorithms, and thus require a matching step between the result and the ground truth regions. Therefore, we can chose more simple measures (see Fig. 4 for the notations): a **completeness metric**:

$$m_5 = \frac{|S_{algorithm} \bigcap S_{ground\,truth}|}{|S_{ground\,truth}|}, \quad m_5 \in [0,1]$$

and a **correctness metric**:

$$m_6 = \frac{|S_{algorithm} \bigcap S_{ground\,truth}|}{|S_{algorithm}|}, \quad m_6 \in [0,1]$$

We can notice that combining $m_5$ and $m_6$, we can compute the Vinet distance. Besides, we can define $m_7$ as a combination of $m_5$ and $m_6$: $m_7 = m_5 \times m_6, \quad m_7 \in [0,1]$, and an overall **quality measure**: $m_8 = m_{3_{left}} \times m_{3_{right}} \times m_7, \quad m_8 \in [0,1]$.

### 5.2.2 Error classification

Hoover et al. proposed an error classification for extracted regions in the scope of image segmentation evaluation. They distinguished correct detection, over- and under-segmentation instances, missed detections and noise [14]. Once more, this classification is better adapted to multiple region matching rather than to a comparison between two regions. Nevertheless, we can notice that the basic values computed to perform this classification are based on boolean operations between pixel sets and correspond to combinations of $m_5$ and $m_6$.

### 5.2.3 Parametric approaches

Finally, concerning parametric approaches, various features of the regions can be computed and compared: surface, perimeter, moments, main axes, etc. Surface and perimeter are also taken into account in the previous measures while moments and main axes (or road axes) may provide interesting additional information.

## 6 PRELIMINARY EXPERIMENTATION

The image database has not been completly delivered and the algorithms are currently being integrated into SENA. We made a preliminary experiment concerning the metrics using two algorithms and one sequence of 224 images. Figure 5 shows the ground truth and the results of both algorithms on the same image. Figure 6 shows the values of the metrics along the image sequence.

Surface-based metrics ($m_5$ and $m_6$) appear far more stable than contour-oriented metrics, which is probably due to the severity of the “buffer method” for small values of the buffer width (12 pixels in our experiment, for $768 \times 576$ pixel images). The peaks in the diagrams indicate particular images for which the algorithms failed. For instance, the right edge determined by algorithm 1 on image 123 (see Fig. 5) presents poor values for $m_1$, $m_2$, $m_4$ and $m_6$. Algorithm 1 faces difficulties on images 81, 89, 90 and 103 as well (see $m_1$ and $m_2$), although $m_4$ indicates that these errors are minor compared to image 123. The end of the sequence presents a greater challenge for the algorithms since the vehicle arrives on a cross-road. As a result, the detectors tend to select a portion of the road which belongs to the intersection and which was not marked by the operator: completeness remains correct while correctness decreases. However, on the rest of the sequence, correctness is better than completeness, which means that part of the road is missed by the algorithms. The road detectors indeed have trouble finding the horizon line, so that the estimated boundaries do not extend to the upper part of the road. A metric that would only consider the lower part of the image would enable us to assess the quality of the algorithm whatever the estimation of the horizon line.

## 7 CONCLUSION

In this paper, we have described the complete methodology and various tools that will be used to assess the quality

Figure 5: Groundtruth (left) and results of algoritm 1 (center) and 2 (right) on image 123.

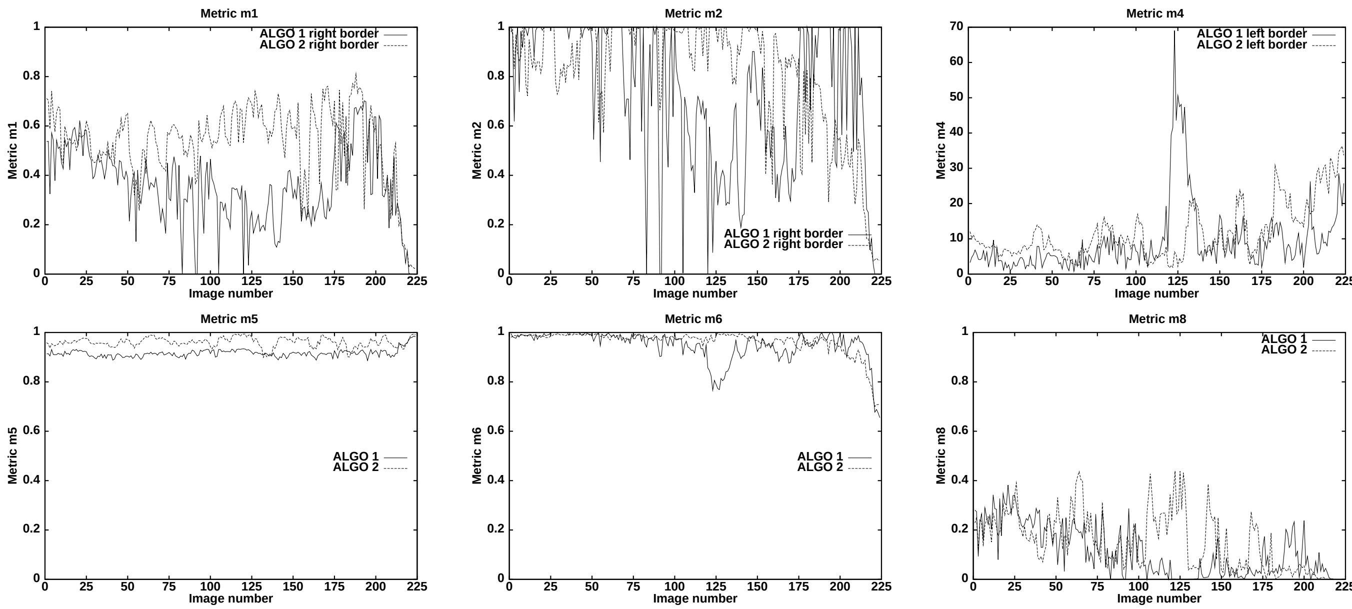


Figure 6: Examples of measures.

of unstructured road edges extraction algorithms. Within the next months, the image database should be completed and all the algorithms will be integrated into the SENA platform. This will allow us to apply our methodology to the whole data and compare the different edge detection techniques. Henceforth, this work offers many perspectives:

- Besides road edge detectors, we plan to apply our methodology to the evaluation of other vision-based algorithms which aim at enhancing the navigation capabilities of autonomous ground vehicles. Among them, we have selected beacon and vehicle tracking as well as image segmentation. The algorithms which we plan to test belong to three French laboratories: Laboratoire des Sciences et Matériaux pour l'Electronique et l'Automatique (LASMEA), Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS) and our laboratory.

- So far, we have defined six different metrics for the automatic assessment of edge detectors. However, we may come to modify these metrics if it turns out that they do not account for some qualitative phenomena observed by the operator during the evaluation. Indeed, Ropert and Capolunghi underline the necessity of a good correlation between the human judgement and the behavior of the metric [6].

- To go further, we could even use a specific methodology for choosing the metrics. Ropert et al. proposed such a methodology in the practical case of default detection in gammagraphy images of welded metal plates [4]. Letournel described a more sophisticated protocole in the field of aerial images interpretation [18]. She performed a statistical analysis in order to detect a relationship between objective metrics (given by mathematical formulas) and subjective metrics given by a human judgement (manual mark-

ings). Such an analysis would definitely be worth trying in the scope of our project.

- Among other metrics that could be tested, we can imagine measures which would be more oriented towards the specific task to be performed by the vehicle, such as the metric described by DeMenthon [8].
- Finally, it seems interesting to introduce metrics that would allow us to characterize more accurately the difficulty of the test images (signal to noise ratio or more sophisticated metrics such as the ones proposed by Kluge [17]). Such metrics should help us to build a more representative video database for the evaluation.

## References


[1] R. Aufrere, R. Chapuis, and F. Chausse. A fast and robust vision based road following algorithm. In *IEEE Int. Conf. on Intelligent Vehicles*, Oct. 2000.

[2] A. J. Baddeley. An error metric for binary images. In *Robust Computer Vision: Quality of Vision Algorithms*, pages 59–78. W. Forstner and S. Ruwiedel, Mar. 1992.

[3] S. Beucher and C.-H. Demarty. Color segmentation algorithm using an HLS transformation. In *Mathematical Morphology and its Applications to Image and Signal Processing, Proceedings ISMM'98*. Kluwer, June 1998.

[4] J. Blanc-Talon and V. Ropert. A software architecture for algorithmic performance evaluation of vision systems. In *Proc. of the $3^{rd}$ Workshop on European Scientific and Industrial Collaboration*, June 2001.

[5] A. Broggi, M. Bertozzi, A. Fascioli, and G. Conte. *Automatic Vehicle Guidance: the experience of the* ARGO *autonomous vehicle*. World Scientific Publishing, 1999.

[6] R. Capolunghi and V. Ropert. L'évaluation empirique en traitement d'images. Rapport technique DGA/DCE/CTA 2000 R 022, March 2000.

[7] P. Charbonnier, P. Nicolle, Y. Guillard, and J. Charrier. Road boundaries detection using color saturation. In *IX European Signal Processing Conf.*, 1998.

[8] D. DeMenthon. Inverse perspective of a road by local image matches and global 3D optimization. Technical Report CAR-TR-210, Univ. of Maryland, 1986.

[9] J. Everson, E. Kopala, L. Lazofson, H. Choe, and D. Pomerleau. Sensor performance and weather effects modeling for intelligent transportation systems applications. In *Intelligent Vehicle Highway Systems (Proc. SPIE vol. 2344)*, pages 118–128, 1994.

[10] A. Guiducci. Parametric model of the perspective projection of a road with applications to lane keeping and 3D road reconstruction. *Computer Vision and Image Understanding*, 73(3), Mar. 1999.

[11] R.M. Haralick. Propagating covariances in computer vision. In *Int. Conf. on Pattern Recognition*, Sept. 1994.

[12] M. D. Heath. A robust visual method for assessing the relative performance of edge detection algorithms. Master thesis, Univ. of South Florida, Dec. 1996.

[13] M. Herman, S. Szabo, K. Murphy, D. Coombs, T.-H. Hong, H. Scott, N. Dagalkis, K. Goodwin, and J. Albus. Recomandations for performance evaluation of unmanned ground vehicle technologies. Technical Report NISTIR-5244, NIST, Aug. 1993.

[14] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.

[15] Q. Huang and B. Dom. Quantitative methods of evaluating image segmentation. In *ICIP'95*, 1995.

[16] T. Jain and R. Binford. Ignorance, myopa and naiveté in computer vision systems. *CVGIP: Image Understanding*, 53:112–117, 1991.

[17] K. C. Kluge. Performance evaluation of vision-based lane sensing: some preliminary tools, metrics and results. In *IEEE Conf. on Intelligent Transportation Systems*, 1997.

[18] V. Letournel. *Evaluation d algorithmes d interprétation pour le renseignement d'origine image*. PhD thesis, DGA/Centre Technique d'Arcueil, 2002.

[19] D. Luzeaux and A. Dalgalarrondo. Assessment of image processing algorithms as the keystone of autonomous robot control architectures. In *Imaging and Vision Systems: Theory, Assessment and Applications*. NOVA Science Books, 2001.

[20] G. Martinoty. Suivi de routes et détection d'intersections. Rapport de stage de DEA IARFA, Univ. Paris VII, Aug. 2001.

[21] W. K. Pratt. *Digital Image Processing*. Wiley, John & Sons, New-York, 1978.

[22] C. Rouquet, F. Chausse, R. Chapuis, and P. Bonton. Segmentation non supervisée d'images de scènes routières : une approche multi-critères. *Traitement du Signal*, 13:197–208, 1996.

[23] F. Sandt and T. Skordas. Color vision for road detection system. In *Session LVIII - Progress in Picture Processing*. NATO Advanced Study Institute, Aug.-Sept. 1992.

[24] R. N. Strickland and D. K. Cheng. Adaptable edge quality metrics. *Optical Engineering*, 32, 1993.

[25] C. Wiedmann, C. Heipke, H. Mayer, and O. Jamet. Empirical evaluation of automatically extracted road axes. In *Empirical Evaluation Techniques in Computer Vision*. IEEE Computer Society Press, 1998.